BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e36187
doi: 10.3897/biss.3.36187

OPEN ACCESS

Conference Abstract

Semantic Annotation of Botanical Collection Data

Dominik Röpert‡, Fabian Reimeier‡, Jörg Holetschek‡, Anton Güntsch‡

‡ Botanic Garden and Botanical Museum Berlin, Berlin, Germany

Corresponding author: Dominik Röpert (d.roepert@bgbm.org)

Received: 14 May 2019 | Published: 13 Jun 2019

Citation: Röpert D, Reimeier F, Holetschek J, Güntsch A (2019) Semantic Annotation of Botanical Collection Data. Biodiversity Information Science and Standards 3: e36187. https://doi.org/10.3897/biss.3.36187

Abstract

Herbarium specimens have been digitized at the Botanical Garden and Botanical Museum, Berlin (BGBM) since the year 2000. As part of the digitization process, specimen data have been recorded manually for specific basic data elements. Additional elements were usually added later based on the digital images.

During the last twenty years, data were transcribed exactly as they were written on the labels, a widely used procedure in European herbaria. This approach led to a large number of orthographic variations especially with regard to person and place names.

To improve interoperability between records within our own collection database and across collection databases provided by the community, we have started to enrich our metadata with Linked Open Data (LOD)-based links to semantic resources starting with collectors and geographic entities. Preferred resources for semantic enrichment (e.g., WikiData, GeoNames) have been agreed on by members of the Consortium of European Taxonomic Facilities (CETAF) in order to exploit the potential of semantically enriched collection data in the best possible way.

To be able to annotate many collection records in a relatively short time, priority was given to concepts (e.g., specific collector names) that occur on many specimen labels and that have an existing and easy-to-find semantic representation in an external resource. With this approach, we were able to annotate 52,000 specimen records in just a few weeks of working time of a student assistant.

© Röpert D et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

The integration of our semantic annotation workflows with other data integration, cleaning, and import processes at the BGBM is carried out using an OpenRefine-based platform with specific extensions for services and functions related to label transcription activities (Kirchhoff et al. 2018).

Our semantically enriched collection data will contribute to a “Botany Pilot,” which is presently being developed by member organizations of CETAF to demonstrate the potential of Linked Open Collection Data and their integration with existing semantic resources.

Keywords

Linked Open Data, LOD, semantic web, Wikidata

Presenting author

Dominik Röpert

Presented at

Biodiversity_Next 2019

References

- Kirchhoff A, Bügel U, Santamaria E, Reimeier F, Röpert D, Tebbje A, Güntsch A, Chaves F, Steinke K, Berendsohn W (2018) Toward a service-based workflow for automated information extraction from herbarium specimens. Database 2018: bay103. https://doi.org/10.1093/database/bay103